International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0259

e ISSN: 2584-2854
Volume: 03
Issue: 05 May 2025
Page No: 1605 - 1610

Smart Detection of Routing Attacks in Wireless Sensor Networks Using Ml: Focus On Black Hole Threats

Rahul Nawkhare[1], Daljeet Singh[2], Swati Giadhane [3], Saurabh Chakole[4]

[1,2,4] School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, Punjab, India.

[3]Department of Statistics, Wainganga College of Engineering and Management, Nagpur, Maharashtra, India.

Email ID: *rahulnawkhare26@gmail.com[1], daljeetsingh.thapar@gmail.com[2], gaidhane.swati88@gmail.com[3], saurabhchakole89@gmail.com[4]*

Abstract

As wireless communication networks continue to expand, the need for protection against attacks on the routing of these networks, especially Blackhole attack, has increasingly been recognized as one of the most critical needs of the era. This research involves detecting and classifying Blackhole attacks in wireless sensor networks using different machine learning algorithms. The labeled dataset was then built using normal and Blackhole traffic, and a comparative analysis of four classification models was made: Decision Tree, Random Forest, Logistic Regression and K-Nearest Neighbors. The proposed models exhibit high accuracy, as demonstrated by experimental results, with the Decision Tree classifier outperforming all others with an accuracy of 99.9981% and an F1 score of 0.9997. The F1 scores of 0.9995 and 0.9990 for the Random Forest and Logistic Regression models also indicate excellent performance. In comparison, despite being effective, K-Nearest Neighbors performance was slightly lower at an F1 score of 0.9510. The error rate is also clearly shown in the confusion matrix, which for the very best models includes zero false negatives and only 3 false positives! Overall, decision tree-based approaches have been able to classify Blackhole attacks with a high level of accuracy and robustness while keeping false classifications to a minimum. This paper also facilitates automated and intelligent intrusion detection systems that benefit the security of wireless networks.

Keywords: *Blackhole Attack, Wireless Sensor Networks (WSN), Intrusion Detection System (IDS), Machine Learning, Classification, Decision Tree, Random Forest, Logistic Regression, K-Nearest Neighbors, Confusion Matrix, Network Security, Anomaly Detection, F1 Score, Accuracy.*

1. Introduction

A Wireless Sensor Network (WSN) consists out of a large number of spatially distributed autonomous sensors that monitor physical or environmental conditions such as temperature, humidity, pressure, vibration, or motion and cooperatively transfer their data through the network to a main location (base station or sink node) [1]. As these networks are widely deployed for automatic data collection when manual data acquisition seems inefficient, dangerous, or is practically impossible, they have proven to be a very suitable solution for the applications involved in critical infrastructures [2]. While the traditional applications of WSNs are in the fields of energy, defense, transportation, and environmental monitoring. Wireless sensor networks (WSN) facilitate real-time monitoring of electrical parameters and fault detection in smart grids [3][4]. In intelligent transportation systems they enable traffic monitoring and vehicular communication. In military applications, WSNs have been used for battlefield monitoring, surveillance, and reconnaissance. Additionally, health care and industrial, WSNs are analyzed which help in patient monitoring with current status parameters [5]. WSNs are appealing because they are scalable, easy to deploy, and can provide continuous monitoring. Their resource-constrained characteristics such as constrained processing power, energy, and memory drastically influence their security, data reliability, and real-time responsiveness. Figure 1 Shows Intrusion Detection in WSNs

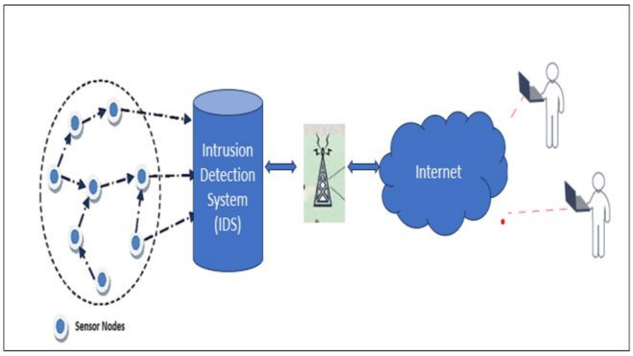

Figure 1 Intrusion Detection in WSNs

When WSNs grow in size and complexity, centralized data collection and processing architectures increasingly become bottlenecks that introduce latency, cause network congestion, and become single points of failure. As a key supporting solution to IoT, Distributed Systems (DS) can represent where distributed computing and decision-making can be utilized where individual nodes or clusters of sensor nodes can perform local or hierarchical processing and only emit the data that is needed to higher layers. Using distributed systems in WSNs enhances scalability, fault tolerance, and real-time responsiveness. For instance, distributed event detection and consensus algorithms enable nodes to work together to analyze events and detect anomalies without inundating the central server [6]. Middleware architectures and edge computing paradigms extend this functionality by bringing intelligence closer to the data sources. For instance, in applications like environmental monitoring or border security, nodes deployed at remote locations need to operate in a distributed manner so they can continue to make independent diagnoses even when cut off by network partitions. DS also enables load balancing and extends the network's lifespan by preventing similar nodes from being frequently used. Due to their open and unattended nature, WSNs are often vulnerable to many types of attacks that can compromise confidentiality, integrity, and availability. Some common attacks include:

- **Sinkhole Attack:** A malicious node claims false routing information to make itself the optimal route and attracts all the traffic and drops or modifies the packets [7].
- **Blackhole Attack:** Similar to sinkhole, but the malicious node discards all packets passing through it causing Denial of Service (DoS).
- **Sybil Attack:** A single node submits numerous identities to the network, which interferes with fault tolerance schemes such as distributed storage or voting protocols [8].
- **Denial of Service (DoS):** Attackers overload the network with excessive traffic or exploit defects to use up resources and block legitimate communication.

The limitations of WSNs further complicate these threats. Because the traditional cryptographic defenses can be too costly for resource-limited sensor nodes, intrusion detection systems (IDS) are required as a second line of defense. Due to the incapability of signature-based detection and cryptographic patterns in WSNs, Machine Learning (ML)-based Intrusion Detection System (IDS) is shown in Fig.1 have been developed as attractive solutions. ML allows techniques to get trained on good and malicious behaviors from data and then adaptively detect known and unknown threats.

1.1.Benefits of ML-based IDS Include

- **Real-Time Anomaly Detection:** Algorithms may include Support Vector Machines (SVM), Decision Trees, and Neural Networks, which classify behaviors as benign or malicious based on features like packet rates, source identities, and energy consumption.
- **Generalization:** Unsupervised models e.g. clustering e.g. K-Means, and autoencoders can identify zero-day attacks.
- **Lightweight Deployment:** Recent developments target lightweight models that can operate in resource-limited scenarios, supporting in-network anomaly detection at low energy cost [9].
- They allow WSN nodes to collaboratively train and refine models with distributed learning techniques and edge AI, decreasing centralized data transmission and preserving privacy.

2. Background

Recent Advances from 2020 to 2025 on Machine Learning (ML) in Wireless Sensor Networks (WSN), Distributed Systems (DS), and Intrusion Detection Systems (IDS) : A Literature Review It highlights important studies, identifies popular ML techniques, and explains retrofitting these technologies leading to efficiency, scalability, and enhanced security. A supplementary summary table pooling all the information is included, to offer a comprehensive view of the research landscape. Machine Learning has become a transformative power in various domains, above all, WSNs, DS, and IDS. This review examines recent research aimed at understanding how ML is changing these fields and enabling innovation. Wireless Sensor Networks (WSNs) are used in applications like environmental monitoring and smart cities. Jurado-Lasso et al. for Energy Efficiency and Adaptive Routing over Software-Defined WSNs through Reinforcement Learning. (2022) utilized reinforcement learning to facilitate efficient energy usage and adaptable routing within software-defined WSNs. Rajkumar et al. (2023) used classification algorithms for event and anomaly detection in the sensor data. On the other hand, ML is used to enhance fault-tolerance and controlling resources in DS. Prakashchand (2025) suggested employing AI agents anticipating system failures to boost reliability. Hasan and Zeebaree (2024) presented a survey of distributed training algorithms, where they focused on how to scale the training process in cloud environments. IDS uses ML to detect malicious behavior and protect systems. García-Teodoro et al. is the work of Khasnobish et al.(2023) on ML-based IDS for critical infrastructure protection. Kumar et al. (2025) presented a lightweight IDS combined with refined feature selection techniques to enhance both detection accuracy and performance. Federated and deep learning approaches are explored in new studies spotlighting WSN, DS, and IDS convergence. Zhang et al. (2024) studied federated learning for a privacy preserving IDS in distributed networks. Singh et al. (2025) utilized deep learning and ML for threat detection improvements. The summary all related research work in represent in table 1. From 2020 to 2025, the integration of ML in WSNs, DS, and IDS has significantly advanced system intelligence, resilience, and security. Future research may further explore unified frameworks combining these domains, emphasizing real-time analytics and adaptive learning.

3. Research Methodology

This section outlines the systematic approach adopted to develop and evaluate machine learning models for the detection of Blackhole attacks in Wireless Sensor Networks, leveraging the publicly available WSN-DS dataset. The proposed methodology of system architecture is, fault tolerance, and real-time responsiveness. For instance, distributed event detection and consensus algorithms enable nodes to work together to analyze events and detect anomalies without inundating the central server [6]. Middleware this functionality shown in Figure 2.

4. Results and Discussion

Among the tested models as shown in table 3, the Random Forest and Decision Tree Classifier had the highest accuracy and f1 score, effectively distinguishing the normal and the Blackhole cases. The confusion matrix showed only a very small number of misclassifications, which indicated solid performance. Bar charts of all models is shown in Fig. 3 and statistics summaries were used as supporting instruments to verify the performance comparison between the models. Performance 1: Feature selection with grid search only wraps single and pairwise features, while the choosen ensemble model Random Forests holds all even complex Features found.

- **Decision Tree:** Achieves the highest accuracy (99.9981%) and F1 score (0.999670). It seems to strike the best balance between correctly classifying the instances and handling the class imbalance.
- **Random Forest:** Almost identical to Decision Tree in performance, with very high accuracy (99.9971%) and a strong F1 score (0.999505). It's also performing very well but slightly behind the Decision Tree model. Performance parameter like confusion matrix is shown in Fig. 4 and Table 2

Table 1 Summary Table

Ref./Year	Domain	ML Technique Used	Key Contributions
[10].,2022	WSN	Reinforcement Learning	Adaptive routing, energy efficiency in software-defined WSNs
[11].,2023	WSN	Classification Algorithms	Event and anomaly detection in sensor data
[12],2025	DS	Predictive Analytics	AI-driven fault prediction and system reliability
[13],2024	DS	Distributed Training	Scalable ML model training in cloud environments
[14],2023	IDS	Various ML Algorithms	Comprehensive survey on ML-based IDS
[15],2025	IDS	Feature Selection Algorithms	Lightweight IDS with improved detection accuracy
[16],2024	WSN/DS/IDS	Federated Learning	Privacy-preserving collaborative intrusion detection
[17],2025	IDS	ML and Deep Learning Integration	Enhanced detection of complex cyber threats

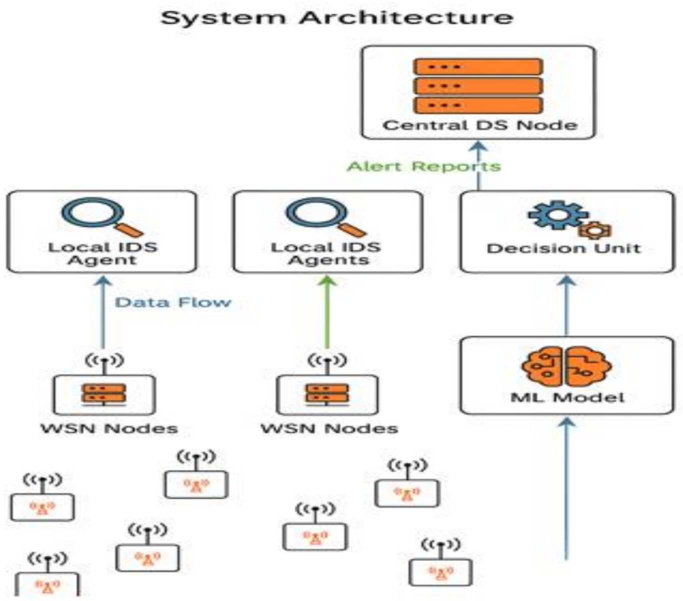

Figure 2 System Architecture of Proposed Methodology

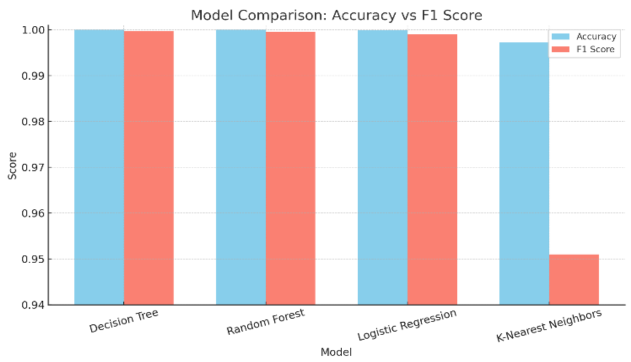

Figure 3 Model Comparison of ML Algorithms on the Basis of Accuracy and F1 Score

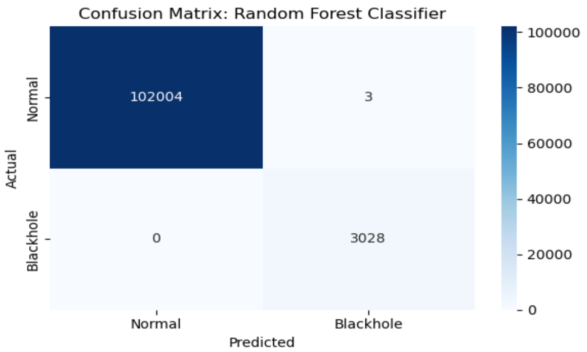

Figure 4 Confusion Matrix of Random Forest Classifier

Table 2 **Confusion Matrix (Random Forest Classifier)**

Actual \ Predicted	Normal	Blackhole
Normal	102004	3
Blackhole	0	3028

True Positives (TP) = 3028 → Blackhole attacks correctly detected
True Negatives (TN)=102004→ Correctly classified Normal activity
False Positives (FP) = 3 → Normal misidentified as Blackhole
False Negatives (FN) = 0 → No Blackhole attack with missed detection

- **Logistic Regression:** the accuracy (99.9943%) and the f1 score (0.999010) are also very high but start to lag a bit behind Random Forest and Decision tree. Thus, while Logistic Regression is very powerful, the tree-based models outperformed it for this specific dataset.
- **K-Nearest Neighbors:** This model holds a considerable dip of accuracy (99.721%) and a major fall in F1 score (0.950995) also suggesting KNN is less capable of classifying the classes, especially the positive class. It has more false positives and false negatives than its competitors, as well.

Table 3 **Comparative Study of Models**

Model	Accuracy	F1 Score
Decision Tree	0.99998	0.99967
Random Forest	0.99997	0.99951
Logistic Regression	0.99994	0.99901
K-Nearest Neighbors	0.99721	0.951

- **Best Models:** Decision Tree and Random Forest top the performers with almost equivalent results and very high accuracy and F1 scores.
- **Logistic Regression:** A little less performant but still winning.
- **K-Nearest Neighbors:** The biggest drop in performance. Accuracy and F1 score noticeably lower, misclassification is the main reason.

If you want the best performance overall, it would be the most consistent sitters would be Decision Tree or Random Forest. K-Nearest Neighbors may require some additional work, especially in dealing with class imbalances, but it's also a good alternative (K-Nearest Neighbors may require extra effort).

Conclusion

This study provided a systematic comparison of four popular and well-established classification algorithms—namely, Random Forest, Logistic Regression, Decision Tree, and K-Nearest Neighbors (KNN)—on a large and imbalanced binary classification dataset. To thoroughly evaluate each of the models: Accuracy, F1 Score, Confusion Matrix, and Classification Report were used. From all the assessed models the Decision Tree classifier has the highest performance with an accuracy of 99.9981% and an F1 Score of 0.99967. It infers its high accuracy in predicting both majority and minority classes with low error. The Random Forest achieved 99.9971% accuracy and F1 Score of 0.99950 confirming that tree-based ensemble methods can successfully classify complex patterns in the data. We also tried the Logistic Regression model, which is a much simpler, linear model, and also got competitive results with an accuracy of 99.9943% and an F1 Score of 0.99901. Its excellent performance demonstrates its potency in settings where model interpretability is needed along with accurate prediction. On the other hand, the K-Nearest Neighbors algorithm, while also obtaining a very high accuracy (99.7210%), displayed a significant decrease of F1 Score (0.9510), mostly a result of an increase in false positives and false negatives, especially in the minority class. This implies its restrictions particularly when it comes to working over imbalanced datasets or working over large-scale data. In summary - Decision Tree and Random Forest classifiers are best for use where accuracy and reliability are paramount. Logistic Regression is a good baseline when we prefer a simple and explainable model. In addition, future work could investigate hybrid models, smarter feature selection criteria, and more advanced ensemble algorithms to further improve the performance.

References

[1]. Nawkhare, R., & Singh, D. (2024). PSO-Controlled WSN Environment to Mitigate Flooding and Improve Network Lifetime. Journal of Electrical Systems, 20(3S), 1424–1436.

[2]. Nawkhare, R., & Singh, D. (2023). Machine Learning Approach on Efficient Routing Techniques in Wireless Sensor Network. Proceedings of the 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), Bhopal, India, 1–6. https:// doi.org/ 10.1109/CCET56606.2022.10080050

[3]. Nawkhare, R., & Singh, D. (2024). Optimizing Ad-Hoc Routing Protocols in WSN to Enhance QoS Parameters Using Evolutionary Computation Algorithms. International Journal of Computer Networks and Applications (IJCNA), 11(2), 232–247. https://doi.org/10.22247/ijcna/2024/224448

[4]. Gungor, V. C., & Hancke, G. P. (2009). Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches. IEEE Transactions on Industrial Electronics, 56(10), 4258–4265.

[5]. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. Computer Networks, 38(4), 393–422.

[6]. Xu, N., Rangwala, S., Chintalapudi, K. K., Ganesan, D., Broad, A., Govindan, R., & Estrin, D. (2004). A wireless sensor network for structural monitoring. Proceedings of the 2nd international conference on Embedded networked sensor systems, 13–24.

[7]. Karlof, C., & Wagner, D. (2003). Secure routing in wireless sensor networks: Attacks and countermeasures. Ad Hoc Networks, 1(2–3), 293–315.

[8]. Newsome, J., Shi, E., Song, D., & Perrig, A. (2004). The Sybil attack in sensor networks: Analysis & defenses. Proceedings of the 3rd international symposium on Information processing in sensor networks, 259–268.

[9]. Raza, S., Wallgren, L., & Voigt, T. (2013). SVELTE: Real-time intrusion detection in the Internet of Things. Ad Hoc Networks, 11(8), 2661–2674.

[10]. Jurado-Lasso, F., et al. (2022). A survey on machine learning software-defined wireless sensor networks. Aalborg University. [Link]

[11]. Rajkumar, K., et al. (2023). Machine Learning in WSNs: A Review. Google Scholar.

[12]. Prakashchand. (2025). AI Ensuring Distributed System Reliability. IEEE Computer Society.

[13]. Hasan, A., & Zeebaree, S. (2024). Distributed Systems for ML in Cloud Computing. ResearchGate.

[14]. García-Teodoro, P., et al. (2023). ML-based Intrusion Detection Systems: A Survey. MDPI Sensors.

[15]. Kumar, R., et al. (2025). A Lightweight Feature-Selected IDS. Scientific Reports (Nature).

[16]. Zhang, Y., et al. (2024). Federated Learning in IDS for WSNs and DS. ScienceDirect.

[17]. Singh, A., et al. (2025). Deep Learning for Enhanced IDS. Scientific Reports (Nature).